DOCUMENT RESUME

ED 333 042 TM 016 582

AUTHOR Natriello, Gary; And Others
TITLE Creating More Responsive Student Evaluation Systems for Disadvantaged Students. Report No. 15.
INSTITUTION Center for Research on Effective Schooling for Disadvantaged Students, Baltimore, MD.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 91
CONTRACT RI17R90002
NOTE 18p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Disadvantaged; *Educational Assessment; Elementary School Students; Elementary Secondary Education; Evaluation Problems; Guidelines; Secondary School Students; *Student Evaluation; Testing Problems
IDENTIFIERS *Performance Based Evaluation; *Responsive Evaluation

ABSTRACT

This paper examines the evaluation of student performance (SP) in schools in order to develop a set of strategies for creating more responsive systems for its evaluation, focusing on disadvantaged students. Problems involved in the evaluation of the performance of disadvantaged students are examined, including: evaluation purposes (direction, motivation, certification, and selection) that are in conflict, task assignments that are overly routine, criteria that are unclear and narrow, standards that deny disadvantaged students the feeling of competence, testing practices that provide limited information on SP, appraisals that limit teacher discretion, feedback that is too limited, and the absence of planning for the improvement of SP. Elements of a more responsive evaluation system for at-risk students include: attention to the purposes of direction and motivation; consideration of tasks prior to developing the evaluation process; broad criteria tapping multiple abilities, processes, and non-academic domains; standards that are challenging yet attainable; a broader range of information and more varied collection techniques; more efficient methods of collecting information on SP; appraisals that involve teacher discretion and broader teacher participation; more detailed feedback; new ways of presenting certification/selection information to students and new ways of presenting direction/motivation to others; and improvement plans that offer opportunities and resources. A 59-item list of references is included. (RLC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED333042

CDS

THE JOHNS HOPKINS UNIVERSITY

Creating More Responsive Student Evaluation Systems for Disadvantaged Students

Gary Natriello, Aaron M. Pallas

and

Carolyn Riehl

Report No. 15

April 1991

CENTER FOR RESEARCH ON EFFECTIVE SCHOOLING FOR DISADVANTAGED STUDENTS

TM016582

BEST COPY AVAILABLE

National Advisory Panel

Beatriz Arias, College of Education, Arizona State University

Mary Frances Berry, Department of History, University of Pennsylvania

Anthony S. Bryk, Department of Education, University of Chicago

Michael Charleston, Educational Policy Studies, Pennsylvania State University

Constance E. Clayton, Superintendent, Philadelphia Public Schools

Edmund Gordon (Chair), Department of Psychology, Yale University

Ronald D. Henderson, National Education Association

Vinetta Jones, Dean, Education and Urban Affairs, Morgan State University

Arturo Madrid, Director, Tomas Rivera Center

Hernan LaFontaine, Superintendent, Hartford Public Schools

Peter Stanley, Director, Education and Culture Program, Ford Foundation

William Julius Wilson, Department of Sociology, University of Chicago

Center Liaison

Harold Himmelfarb, Office of Educational Research and Improvement

Creating More Responsive Student Evaluation Systems for Disadvantaged Students

Gary Natriello, Aaron M. Pallas and Carolyn Riehl

Teachers College, Columbia University

Grant No. R117 R90002

Report No. 15

March 1991

Published by the Center for Research on Effective Schooling for Disadvantaged Students, supported as a national research and development center by funds from the Office of Educational Research and Improvement, U.S. Department of Education. The opinions expressed in this publication do not necessarily reflect the position or policy of the OERI, and no official endorsement should be inferred.

Center for Research on Effective Schooling for Disadvantaged Students
The Johns Hopkins University
3505 North Charles Street
Baltimore, Maryland 21218

The Center

The mission of the Center for Research on Effective Schooling for Disadvantaged Students (CDS) is to significantly improve the education of disadvantaged students at each level of schooling through new knowledge and practices produced by thorough scientific study and evaluation. The Center conducts its research in four program areas: The Early and Elementary Education Program, The Middle Grades and High Schools Program, the Language Minority Program, and the School, Family, and Community Connections Program.

The Early and Elementary Education Program

This program is working to develop, evaluate, and disseminate instructional programs capable of bringing disadvantaged students to high levels of achievement, particularly in the fundamental areas of reading, writing, and mathematics. The goal is to expand the range of effective alternatives which schools may use under Chapter 1 and other compensatory education funding and to study issues of direct relevance to federal, state, and local policy on education of disadvantaged students.

The Middle Grades and High Schools Program

This program is conducting research syntheses, survey analyses, and field studies in middle and high schools. The three types of projects move from basic research to useful practice. Syntheses compile and analyze existing knowledge about effective education of disadvantaged students. Survey analyses identify and describe current programs, practices, and trends in middle and high schools, and allow studies of their effects. Field studies are conducted in collaboration with school staffs to develop and evaluate effective programs and practices.

The Language Minority Program

This program represents a collaborative effort. The University of California at Santa Barbara is focusing on the education of Mexican-American students in California and Texas; studies of dropout among children of recent immigrants are being conducted in San Diego and Miami by Johns Hopkins, and evaluations of learning strategies in schools serving Navajo, Cherokee, and Lumbee Indians are being conducted by the University of Northern Arizona. The goal of the program is to identify, develop, and evaluate effective programs for disadvantaged Hispanic, American Indian, Southeast Asian, and other language minority children.

The School, Family, and Community Connections Program

This program is focusing on the key connections between schools and families and between schools and communities to build better educational programs for disadvantaged children and youth. Initial work is seeking to provide a research base concerning the most effective ways for schools to interact with and assist parents of disadvantaged students and interact with the community to produce effective community involvement.

Abstract

This paper examines the evaluation of student performance in schools in order to develop a set of strategies for creating more responsive systems for the evaluation of student performance. The paper examines problems involved in evaluation of the performance of disadvantaged students, including problems in the purposes of evaluation, task assignment, criteria, standards, testing, appraisal, feedback, and planning for improvement. A set of guidelines is derived for creating evaluation systems that would be more responsive to disadvantaged students.

Introduction

This paper examines the evaluation of student performance in schools in order to develop a set of strategies for creating more responsive systems for the evaluation of student performance. The concept of responsive evaluation systems derives from a more general concept of responsive schools. Responsive schools are those in which student performance and participation are met with timely and appropriate reactions from adults (Natriello, McDill, and Pallas, 1990; McPartland and McDill, 1977). In responsive schools, staff recognize and reward good student performance and identify and penalize poor student performance. Moreover, school staff identify the larger causes of poor student performance and develop remedies, both interpersonal and structural, to create conditions under which students can succeed.

Evaluation processes are key elements in schools (Natriello, 1987). Student performance and behavior are routinely monitored and evaluated. Indeed, many of the activities of teachers, from formal testing to informal observations and comments, can be seen as evaluation of student performance in the context of the instructional and extracurricular program of the school. It is through evaluations that the school attempts to influence the development of students. Thus, evaluations and student evaluation systems are a central element in any attempt to create more responsive school environments.

The process of evaluating student performance in schools and classrooms is complex and involves a variety of activities. Each of these elements may present problems for disadvantaged students. Evaluation systems are often characterized by (1) purposes that are in conflict, (2) tasks that tend to be overly routine, (3) criteria that are unclear and narrow, (4) standards that deny disadvantaged students the feeling of competence, (5) tests that provide limited information on student performance, (6) appraisals that limit teacher discretion, (7) feedback that is too limited, and (8) the absence of plans for the improvement of student performance. We consider these and other problems in evaluation systems as a prelude to developing a series of guidelines for creating evaluation systems that would be more responsive to disadvantaged students.

Problems in Evaluation Systems for Disadvantaged Students

Purposes In Conflict

There are at least four purposes for student evaluation systems: direction, motivation, certification, and selection. Direction refers to the goal of providing students with clear and specific information on their performance and how to modify it along the lines desired in the school. Motivation involves securing the cooperation and commitment of those being evaluated to the tasks at hand. Both of these purposes are compatible with the development of responsive evaluation systems. Evaluation activities oriented to these purposes can lead to school and classroom practices that respond to the needs and performance problems of all students, particularly disadvantaged students, who often lack access to external sources of motivation and direction.

However, the other two purposes for evaluation systems, certification and selection, do not necessarily lead to activities that respond to student needs. Certification and selection respond to the needs of others, not to the needs of students. Certification provides assurance to others that a student has attained a certain level of accomplishment or mastery. Selection is the identification of some students for future educational or occupational opportunities. Although certification and selection offer certain benefits to those students who are certified and/or selected, they lead to practices that contribute little to the development of students.

The tension between the purposes of certification and selection and those of direction and motivation are only apparent in the case of the evaluation of youngsters who have trouble succeeding in school. Students with reasonable chances of success profit from the certification and selection functions of evaluations as well as from the direction and motivation functions. For disadvantaged students, who are unlikely to experience success as a result of those practices set in motion by the certification and selection functions of evaluation, the direction and motivation functions become even more important. However, the basic tension between direction and motivation, on the one hand, and certification and selection, on the other, creates problems in design-

ing all aspects of evaluation processes for disadvantaged students.

Assigning Tasks: Between Boredom and Anxiety

Evaluation processes set up situations in which all students are at risk of failure. The level of risk varies with the nature of the tasks on which students are being evaluated. Dornbusch and Scott (1975:80) have suggested that tasks differ in predictability, that is, "the extent to which the performer has knowledge of which path is most likely to lead to success." Furthermore, they demonstrated that when tasks are predictable, performers prefer that they be given the specific procedures to be followed to execute the tasks, but that when tasks are less predictable, performers prefer the autonomy to make their own decisions about how to perform the task.

Predictable tasks accompanied by specific procedures to be followed carry relatively less risk than less predictable tasks. In an analysis of classroom tasks Doyle (1983) classified opinion tasks and certain memory tasks (i.e., those involving the reproduction of small amounts of material) and certain routine tasks (i.e., those requiring relatively simple algorithms) as low in risk, and understanding tasks and other memory tasks (i.e., those involving the reproduction of large amounts of material) and other routine tasks (i.e., those involving complicated procedures) as high in risk.

Research in classrooms has revealed two forces at work to reduce the risk associated with the evaluation process. First, Davis and McNight (1976) and Wilson (1976) showed that students resist attempts to shift from low risk routine or procedural tasks to high risk understanding tasks in classrooms. Moreover, students also attempt to renegotiate assigned tasks so that they are more predictable and entail less risk.

Second, there appears to be a tendency to structure evaluation activities as if the tasks being evaluated are predictable. For example, Armbruster, Stevens, and Rosenshine (1977), in a study of three reading curricula, found that although the texts emphasized comprehension and interpretation skills, the tests solicited factual information from students based on the ability to locate information in the text. Bailey, Brazee, Chiavaroli, Herbeck, Lechner, Lewis, McKittrick, Redwine, Reid, Robinson, and Spear (1988) found that despite an intention to view reading and writing as meaning-centered processes, the evaluation system tended to encourage practice in isolated drill and skill exercises. To counteract this it was necessary to create evaluation and reporting instruments which monitored the extent to which students took risks as readers and writers to engage in learning tasks consistent with the meaning-centered approach.

Studies of typical classrooms that reveal a tendency to avoid less predictable tasks and utilize routine tasks to minimize the risk associated with evaluation processes suggest that these tendencies may be even more severe in classes serving disadvantaged youngsters. Such youngsters are even more likely to experience failure under traditional evaluation processes and so face even greater risks when confronting less routine tasks. However, the retreat from unpredictable tasks to minimize risk and anxiety may leave disadvantaged students in classrooms where the instructional tasks are insufficiently interesting to hold their attention (Oakes, 1985).

Criteria: Focusing Attention or Narrowing Opportunity

Criteria are those properties of tasks that are taken into account when making an evaluation of performance. For example, in assessing student performance in writing, teachers may apply criteria regarding creativity and imagination as well as criteria pertaining to correct grammar and proper usage.

Several issues surrounding the development and use of performance criteria have important implications for the evaluation of disadvantaged students. Studies of the impact of evaluation processes on students have revealed that criteria are often not clearly communicated to students (Natriello, 1982) and that criteria are often not linked to the instructional program (Linn, 1983; Rudman, Kelly, Wanous, Mehrens, Clark, and Porter, 1980). Each of these problems presents difficulties for all students, but the difficulties may be worse for disadvantaged students, who lack family resources to help them understand the demands of their teachers and direct their efforts accordingly.

Efforts to clearly articulate criteria and link them to the instructional process have the potential for improving the evaluation process for disadvantaged youth by focusing attention on the most important aspects of the curriculum. If these efforts at greater articulation and alignment result in a narrowing of the criteria and a specification of tasks to the point where routine tasks dominate the classroom, however, they also have the potential to create additional problems for disadvantaged youth. The narrowing of criteria for the evaluation of performance results in a reduction of opportunities to experience success in the classroom. For

disadvantaged students, who typically have difficulty with academic work, a reduction in opportunities to succeed may have serious negative effects on motivation. Moreover, if classroom tasks become overly routinized, students may perceive them as boring and uninteresting. There are practices and special efforts to broaden the base of criteria for evaluations. As Thorndike (1967) notes:

> In practice, certainly many other considerations than that of pure competence do enter into marks. Such factors enter in as (1) industry and effort -- i.e., completing all assigned work and even doing optional work for "extra credit" (a kind of educational bribe); (2) frequent and active participation in class discussion; (3) neatness in written work and mechanical correctness in such areas as spelling and grammar; and (4) personal agreeableness, attractiveness, cleanliness, and docility. To some extent and by some instructors, certain of these factors would be endorsed as legitimate influences on a mark. Others would more uniformly be accepted as extraneous influences to be minimized as far as possible. (p. 762)

Holmes (1978) notes that other criteria which make it possible for an organization to operate often covertly enter into the evaluation process. These include criteria related to behavior and effort such as politeness, conformity, and perseverance. Salganik (1982) reports on a system in which students were evaluated on three criteria -- achievement, effort, and conduct. In a national survey of secondary school teachers, Natriello and McPartland (1987) found that student effort was "very important" or "extremely important" in the evaluation process of over 70% of the teachers.

Many of the strategies for broadening the set of criteria upon which evaluations are based involve including criteria of a more formative nature (i.e., those factors, such as effort and perseverance, that contribute to ultimate achievement) as well as summative factors. Rudman (1987) discusses this as a strategy of linking assessment with teaching and Broadfoot (1986) sees it as attending to the curriculum function of reporting systems as well as to the communication functions. Parkay (1987) provides some initial evidence that such approaches may be particularly effective for disadvantaged students. He presents evidence indicating that students in remedial English classes in an urban high school demonstrated greater knowledge and more favorable attitudes toward the subject matter after being graded for a 3-week period on just the completion of assignments instead of traditional grading procedures which take into account the quality of the work completed.

Broadening the base of criteria upon which students are evaluated offers increased opportunities for disadvantaged youngsters to succeed in schools. Of course, one strategy is simply expanding the range of the academic performance criteria. One approach to providing such opportunities for success for students who would otherwise experience consistent failure in the classroom involves a restructuring of the tasks of the classroom so that they draw on a wider range of ability dimensions. Such multiple-ability classrooms (Cohen, 1986; Rosenholtz, 1977) attempt to move beyond the narrow range of academic tasks and criteria, most of which rely upon reading skills, so that every student can experience some success. In the multiple-ability classroom, the intention is for all students to find some task at which they can experience a sense of competence.

Standards: Between A Rock and A Hard Place

Criteria are those dimensions of tasks upon which evaluations will be based. Standards are the levels of performance expected on the criteria. Research and commentary on appropriate standards for the evaluation of student performance have focused on three types of standards: those set in reference to the criterion level of a group, those set in reference to some absolute criterion level, and those set in reference to the previous criterion level of an individual (Wise and Newman, 1975; Rheinberg, 1983; Thorndike, 1969). Each of these types of standards has different implications for the evaluation of disadvantaged students.

Terwilliger (1977) links norm-referenced standards to what he terms the pragmatic philosophy, a viewpoint primarily concerned with practical choices and the consequences of such choices. An evaluation system which differentiates among individual students is optimal for identifying the choices available to those who make decisions about admission to future educational and employment opportunities. Thus, norm-referenced standards serve the purpose of selection identified earlier.

Norm-referenced standards have been criticized in many quarters. Bresee (1976) lists a series of problems with such standards:

(1) the conflict between the necessity of producing a normal distribution of grades and the goal of having teachers produce improvement in all students in a class;

(2) the distortion of the curriculum as teachers seek to diversify instructional objectives to produce a range of achievement in a class;

(3) the diversion of student attention from the task at hand to the performance of other students; and

(4) the introduction of false competition because achievement is not really in limited supply.

To these Deutsch (1979) adds: (1) the distortion of the testing process so that tests take the form of contests in which all performers participate under uniform conditions; (2) the lack of rewards created by the artificial scarcity of good grades that is likely to impede the development of students' sense of their own value; and (3) the encouragement of competition which may be counterproductive for tasks requiring cooperation and communication. Much of the criticism of norm-referenced standards has to do with their being ill-suited to the purposes of certification (Glaser, 1963), direction and motivation (Beady and Slavin, 1981).

Absolute or criterion-referenced standards have enjoyed increased support in recent years. They are based on some externally determined level of performance. Terwilliger (1977) associates such standards with the behaviorist perspective. The actual strategy for setting such absolute standards has been challenged by Glass (1978) and Burton (1978), who conclude that they must be set arbitrarily, and by Shepard (1976), who concludes that all current methods of setting absolute standards reduce to some form of norm-referenced standards. Criterion-referenced standards may serve the purpose of certification, but they often fail to provide motivation and direction to students, particularly when decisions about standards are removed from the informed professional opinion of teachers (Burton, 1978).

Individually-referenced or self-referenced standards are based on comparing a student's current performance with some other feature of the student. Terwilliger (1977) points out that current performance may be compared with earlier performance or with a student's ability. Terwilliger (1977) associates self-referenced standards with the humanist view of education, and sees them as a strategy to recognize individual differences, reward effort, and foster motivation.

Disadvantaged students are likely to have limited opportunities to experience success under norm-referenced or criterion-referenced standards. Systems based at least in part on self-referenced standards would offer disadvantaged students some sense of competence. The dilemma, of course, is that systems that rely solely upon self-referenced standards deny students information on how their performance may be viewed by a broader audience. Such systems may fail to challenge students and leave them with misconceptions about their own performance (Natriello and Dornbusch, 1984; McDill, Natriello, and Pallas, 1985).

Collecting Information: Tests and Contests

Collecting information on student performance for the purposes of evaluation is a demanding and time-consuming task. Because it is impractical if not impossible to collect complete information on student performance, the collection of such information involves sampling decisions. The challenge is to insure that the information collected provides a valid and reliable estimate of performance appropriate to the purposes, tasks, criteria, and standards of the evaluation process. A number of analysts have contributed important observations about the relationship between the dominant method of collecting performance information -- testing -- and the purposes, tasks, criteria, and standards for evaluation. Deutsch (1979) argues that the structure of most testing situations is a reflection of the prevailing purpose of evaluation (selection) and the types of standards utilized (norm-referenced). He notes that:

> The social context of most educational measurement is that of a contest in which students are measured primarily in comparison with one another rather than in terms of objective accomplishment. If educational measurement is not mainly in the form of a contest, why are students often asked to reveal their knowledge and skills in carefully regulated test situations designed to be as uniform as possible in time, atmosphere, and conditions for all students? Individuals vary enormously in terms of the amount of time they need and the kind of atmosphere and circumstances that facilitate or hinder their expression of their knowledge and skills; it is only the comparison of students with one another that requires measures of educational achievement that take the form of contests (Deutsch, 1979: 394).

As long as testing is constructed as a contest with a limited number of winners, disadvantaged students are likely to turn out the losers. Deutsch goes on to describe the damaging effects of norm-referenced standards for individual students and advocates an evaluation system that would provide individualized, particularistic feedback to students to foster their development. Thus, his objection to the typi-

cal testing situation is rooted in a rejection of evaluation for the purpose of selection and of standards that are norm-referenced. Of course, both selection and norm-referenced standards are likely to lead to practices inimical to disadvantaged students.

Although much of what has been written on the collection of information on student performance has centered on tests and testing, there is a range of alternative methods of collecting such information. For example, Gaston (1976) suggests monitoring students' unassigned reading in the library and listening to student conversations as they leave the classroom as a way to collect information on student attitudes and behavior. Heller (1978) suggests alternatives to standardized reading tests such as the use of reading materials from popular magazines, fables, and poems. Solo (1977) explains how alternatives such as anecdotal records and collections of students' daily work may be used to provide insight into student performance. Herman and Dorr- Bremme (1984) note a variety of techniques used by teachers to collect information on student performance, including routine class and homework assignments, classroom interaction during question and answer sessions, recitations, discussions, oral reading, problem-solving at the chalkboard, special projects, presentations, and reports.

Such alternative methods of collecting information on performance may provide disadvantaged students with a wider variety of ways to demonstrate their competence. At the very least, they provide those who work with disadvantaged students with a broader base of information for understanding their academic performance and needs.

Appraisals: Sensitivity versus Bias

Appraisal involves comparing the information collected on student performance on assigned tasks with the criteria and standards previously established for those tasks. But appraisal is a more complex process than this description would suggest. As Dornbusch and Scott (1975) observe:

> The application of standards in specific situations is rarely a simple or straightforward procedure. It requires judgment with respect to the comparability of the performance situation and the situations for which the standards are considered applicable ... appraisal is seldom a mechanical procedure. In short, appraisal entails deciding how to interpret a low or high performance score. Accurately appraising a task performance requires knowledge of extenuating circumstances, whether it be the inexperience of the task performer, the lack of facilities, or assistance received from a more skilled coworker. Such information is of critical importance in determining what, if any, message is to be communicated to the performer concerning the quality of his or her task performance (p. 143).

But the exercise of discretion that is so critical to producing an appropriate appraisal is viewed as dangerous in the case of teachers appraising the performance of students (Archer and McCarthy, 1988), particularly disadvantaged students. The fear, of course, is that teachers will be biased against disadvantaged students either by failing to recognize their strengths (e.g., Rosenthal and Jacobson, 1968; Rist, 1970), or failing to challenge them to perform to their capacity (Natriello and Dornbusch, 1984), or both.

This fear has led teachers and those who supervise them to seek to reduce teacher discretion in the evaluation process. This can be done through increased used of standardized tests and through teacher grading practices that are more formulaic and allow less room for judgment. These practices are seen as easier to defend, if they are challenged (Natriello, 1982). However, the dilemma is that such practices limit the exercise of teacher discretion which may be important to developing evaluations that provide sufficient direction and motivation to disadvantaged students. Teachers, those individuals in the best position to be more sensitive to students, their performance, and their condition, may be overly constrained by fears of possible bias. Of course, neither teacher bias nor constrained teacher discretion is beneficial to disadvantaged students.

Full teacher discretion without bias represents the best condition for appraisal of the work of all students, including disadvantaged students. One strategy for achieving such conditions might be to broaden the production of and the audience for teacher appraisals. Currently, most teacher appraisals of student performance are the product of a single teacher and are meant for a single student and perhaps his or her parents. But it is possible to involve more than one teacher in the appraisal process or for teachers to share their appraisals with each other. These processes would preserve and perhaps even increase teacher discretion while providing checks and balances on the possible biases of any single teacher. To encourage teachers to work together on the appraisal of student performance it will be necessary to overcome teachers' habits of not relying on information from other teachers (Dorr-Bremme, 1983).

Feedback: The Economies of Providing Information

Providing feedback on performance evaluations is an important component of the evaluation process. If the results of evaluations are never communicated, there is little reason to expect them to have an effect. The results of evaluations may be communicated to multiple audiences, including the student, parents, school officials, and potential employers (Ahmann and Glock, 1967).

The feedback process can be quite varied and contribute to the multiple purposes of evaluations. Slavin (1978) distinguishes three types of feedback regarding student performance: feedback that tells students where they stand compared to other students, based on norm-referenced standards; feedback that provides students with information on their on-going performance, based on criterion-referenced standards; and feedback that enhances student motivation, based on timely assessments on tasks that are neither too difficult nor too easy. Lissman and Paetziod (1983) distinguish between informative feedback and motivational feedback.

Feedback creates several dilemmas for the evaluation of disadvantaged students. First, different kinds of feedback will often result in quite different effects on students. For example, disadvantaged students may be achieving quite poorly in terms of norms for the broader population of students, but they may be achieving quite well in terms of their previous performance. If students receive information on their performance in relation to the norms they may be disheartened. If they receive information on their performance in relation to their previous performance, they may be misled about their opportunities in the broader society. If they receive both kinds of feedback they may be quite confused as to where they actually stand. Thus, providing the full range of feedback to disadvantaged students is a complex undertaking.

Second, providing feedback is a time-consuming process that is often made manageable by the reduction of information on student performance to common symbols such as grades. Although these symbols may serve the needs of external audiences quite well (e.g., prospective employers), they often do not serve the needs of students for detailed information on performance problems to shape future efforts to achieve. Moreover, shorthand symbols often do not serve the needs of school officials, who must match student needs with available school resources. A grade of "D" given to 2 students may result from very different academic problems requiring different remedies, yet the single grade fails to communicate a clear and detailed meaning tied to a range of possible responses.

Detailed feedback seems to increase students' motivation and performance. Butler and Nisan (1986) found that sixth graders scored higher on both a quantitative task and a qualitative (divergent thinking) task when they received task-related comments on their performance. In contrast, students receiving grades scored high on the quantitative task and low on the qualitative task, and students receiving no feedback scored low on both tasks. Moreover, students receiving written comments expressed greater interest in the tasks than students in the other two groups.

There may be particular advantages to using detailed feedback with disadvantaged students. Maas-de Brouwer (1986) reports on several studies indicating that initially poor performing students profit more from detailed feedback and criterion-centered feedback, and that these effects occur in the academic, social, and affective domains.

The dilemma is that providing detailed feedback requires a great deal of time, yet detailed feedback provided shortly after the performance being evaluated is what students, particularly disadvantaged students, require to plan future efforts accordingly. A number of investigators are developing alternatives to traditional feedback processes. Rudman (1976) advocates the use of reporting devices such as checklists that are more closely related to the mechanisms for recording student performance. Such mechanisms may allow greater efficiency in the feedback process. Ediger (1975) suggests more frequent and more varied mechanisms for reporting students' performance, including telephone and face-to-face conferences with parents. Giannangelo and Lee (1974) and Giannangelo (1975) describe a system of computer-assisted reporting that provides more anecdotal information on student performance. Holtz (1976) presents a reporting method for students' performance in elementary science more clearly related to evaluation criteria.

Natriello (1982) reports on non-traditional methods such as the use of cassette tapes of teachers' comments, open gradebooks that students can examine at any time, and procedures for involving students in tallying their own cumulative scores at various points in the grading period. Walling (1975) discusses five broad categories of reporting techniques -- traditional grades, percentage ratings, checklists of objectives, narrative evaluations, and conferences. Stewart (1975) describes a multi-dimensional reporting system for use in elementary schools.

Investigators and practitioners are also considering the use of student profiles or portfolios, composites of various kinds of information on student performance (Bank & Williams, 1987; Mayfield, 1983; Phoenix Union High School District, 1981). Withers (1986) presents a typology of the various kinds of profiles and Broadfoot (1986) describes one such profile, the record of achievement in use in England and Wales.

Planning for Improvement

In terms of improving student performance, evaluations of student performance that end with the communication of the evaluation may have little impact, though such evaluation processes may serve the purposes of certification and selection. Simply communicating an evaluation is unlikely to provide adequate direction and motivation for students. Thus, it is important to link the evaluation process with efforts to plan and deliver instructional or other services that can help students, especially disadvantaged students, to succeed in school.

In particular, evaluative information should enable schools to identify student needs more accurately and to assign school resources -- instruction, counseling, books and materials, and so on -- where and when they are needed by students. Unfortunately, most schools are not equipped to respond quickly and appropriately on the basis of information about student performance. Student assessment information is used in school planning efforts only in the most rudimentary ways -- for example, in infrequent decisions about student course placement. Even teachers, who work most directly with students, base their planning decisions on factors other than student assessments (Clark & Peterson, 1986; Shavelson & Stern, 1981; Stiggins, Conklin, & Bridgeford, 1986).

Often, school staff plan on the basis of trends in the assessments of previous cohorts of students, because they cannot obtain and process current assessment information quickly enough. When teachers and other school staff do not act on the basis of student assessment information, students and their parents must assume the major burden of developing appropriate responses to evaluative information. This is especially problematic for disadvantaged students who lack the personal, family, and community resources to interpret evaluations and react appropriately to them.

Elements of a More Responsive Evaluation System for At-Risk Students

The literature on the evaluation of students in schools and classrooms provides a number of insights for the design of evaluation systems for at-risk students. The same systems that are designed with disadvantaged students in mind are also likely to be appropriate for all students. These systems would incorporate the following elements.

1. Attention to the Purposes of Direction and Motivation

Because evaluation systems typically encompass multiple purposes, it is useful to consider each of the purposes in the design of the system. Because evaluation systems have most often neglected the purposes of direction and motivation, it is particularly important to consider these purposes when designing evaluation systems for disadvantaged students. This can be done by considering each of the elements of the system in terms of these purposes.

2. Consideration of Tasks Prior to Developing the Evaluation Process

Existing evidence indicates that evaluation processes often lead to the development or interpretation of tasks in ways that minimize the risk experienced by those being evaluated by resorting to routinization. Such routinization, while offering some protection to students from evaluation processes under conditions of uncertainty, also leads to less motivating school experiences. These trends are even more pronounced in classes and programs serving disadvantaged students.

To avoid such a trend it is necessary to consider the nature of the tasks assigned to students to determine how best to include less routine and more engaging tasks, interesting tasks that carry reasonable risks. Considering tasks prior to developing the evaluation process should prevent the evaluation process from driving the selection and definition of tasks. In this way educators can assure an adequate representation of non-routine tasks in the school experiences of disadvantaged students. Some proportion of routine tasks might be included to offer students greater security in terms of being able to secure good evaluations for following the specified courses of action, particularly when such tasks are integral to mastering a certain body of knowledge.

3. Broad Criteria Tapping Multiple Abilities, Processes, and Non-academic Domains

Current evaluation systems have been criticized for focusing on overly narrow conceptions of what is legitimate to learn in school and what kinds of competencies should be assessed. Educators of disadvantaged students should develop broader criteria for the evaluation of students. Such broadening of evaluation criteria can take place in three ways.

First, criteria for academic task performance can be broadened to encompass more areas of competence. It might be particularly important to develop criteria that are not closely related to reading ability. Second, criteria can be enlarged to include the processes that lead to successful task completion as well as the final results of tasks. Devoting attention to process criteria will have both directing and motivating effects as students receive both better directions on how to approach tasks and reinforcement for devoting the effort to specified processes. Third, evaluation criteria can include non-academic areas of student performance and development. Social and affective dimensions of student performance often play a role in their ability to succeed in school, and an evaluation system can both direct and motivate students to appropriate non-academic behavior in ultimate support of their academic performance.

4. Standards that are Challenging yet Attainable

Developing appropriate standards for disadvantaged students is not as simple as the recent school reform movement would suggest. It is not just a question of raising standards to challenge students to put forth greater effort and realize greater achievement. There are at least three kinds of standards that must be considered in evaluating students: absolute standards, relative or norm-referenced standards, and individual or self-referenced standards. Evaluation systems that primarily serve the purpose of selection place emphasis on norm-referenced standards, while those that primarily serve the purpose of certification place emphasis on absolute standards.

Systems of evaluation designed to direct and motivate students should include self-referenced standards which allow students to gain a sense of their own progress. In reality, most evaluation systems contain elements of each of these three types of standards. The difficulty is in deciding how to combine different kinds of standards to develop an evaluation of students. A key to any solution to this situation is to understand clearly which type of standard is being applied to produce any given assessment.

5. A Broader Range of Information and More Varied Collection Techniques

The dominant techniques for collecting information on student performance are limited in scope and in the conditions under which they are employed. Moreover, techniques such as formal testing are better suited to the selection and certification purposes of evaluation than to direction and motivation. Techniques of gathering information on performance under varying conditions must be developed to serve the direction and motivation purposes of evaluation. Particular attention must be devoted to making data collection experiences informative and engaging for disadvantaged students.

One strategy would be to model information-gathering activities after real-life experiences of the students, i.e., collect information under conditions similar to those in which students are likely to find themselves in the community or in the workplace. One implication of such a strategy would be to eliminate assessment situations that force students to perform within artificial time limits. Moreover, such information gathering should extend over a broader domain of performance if we hope to understand the nuances of the performance difficulties experienced by disadvantaged students.

6. More Efficient Methods of Collecting Information on Student Performance

Because collecting information on student performance is labor intensive, any attempt to broaden the range of data and the range of conditions under which they are collected must be accompanied by careful thinking about how to make the collection of such performance information more efficient. Teachers have limited time for evaluation activities, and they must be helped to develop less labor-intensive strategies for assembling information on performance. Another approach would be to involve teacher aides in the collection of the broader range of information on performance and leave the development of the appraisal to the teacher.

7. Appraisals that Involve Teacher Discretion and Broader Teacher Participation

The dominant ways of evaluating students leave too little discretion for teacher participation. This is once again due to the selection and certification purposes of evaluation. If evaluations are to be compared across students or against a common standard, it is necessary to severely constrain the teachers' role in developing and executing the eval-

uation activities. However, teachers are in the best position to structure evaluation activities to reinforce the instructional program and contribute to the direction and motivation purposes of evaluation. Teachers should be encouraged and taught how to develop evaluation processes that support learning and student development. Teachers should be able to "invent" evaluation processes that match the needs of their students.

8. More Detailed Feedback

It is clear that disadvantaged students need and profit from more detailed feedback on their performance. Detailed feedback that is related to the learning tasks serves not only to communicate the result of one performance assessment cycle but also to begin the next round of instruction. Providing students with detailed feedback requires a great deal of time. Teachers should be assisted to develop feedback strategies that result in the most detailed feedback within the time constraints of their positions. However, serious attention to detailed feedback is likely to require greater resources in terms of teacher time, perhaps through alterations in the school schedule.

9. New Ways to Present Certification/ Selection Information to Students and New Ways to Present Direction/Motivation to Others

Evaluation systems often present summary information on student performance to students and outsiders. For disadvantaged students, such information is often a communication of failure. For outsiders such information is often a communication of the failure of disadvantaged students relative to less disadvantaged students. In each case the communication presents problems. Disadvantaged students need to understand how they compare to others and to established standards, but they need to under stand this in the context of their own development. Disadvantaged students can be making progress and still receive consistently negative information on their standing in relation to others or established standards. They often lose sight of the former in the presentation of the latter.

Teachers, administrators, counselors, and parents also need to receive information on student performance in broader terms than just the relative comparisons if they are to move beyond the stereotypical assessment of disadvantaged students to developing creative solutions that take advantage of more complex patterns of student development.

10. Improvement Plans that Offer Opportunities and Resources

It does little good to communicate the results of an evaluation to students if we are not prepared to supply them with the resources they need to improve their performance. This is particularly true for disadvantaged students, for whom the evaluative information is often negative. Perhaps we should adopt a rule of thumb that the only evaluative information that may be communicated is that which is accompanied by a plan for improvement complete with the appropriate learning resources. Evaluations would conclude not with assessments of performance but with assessment/opportunity packages that are conveyed to students.

Incorporating these ten elements into a system for evaluating student performance would produce evaluation systems that differ dramatically from those presently operating in most schools. These elements constitute a strategy for taking more seriously the direction and motivation functions of evaluation systems. As such, they should lead to more responsive schools for all students.

References

Ahmann, J., & Glock, M. D. (1967). *Evaluating pupil growth: Principles of tests and measurements.* (4th edition). Boston: Allyn and Bacon.

Archer, J. & McCarthy, B. (1988). Personal biases in student assessment. *Educational Research, 30,* 142-145.

Armbruster, B. B., Stevens, R. J., & Rosenshine, B. (1977). *Analyzing content coverage and emphasis: A study of three curricula and two tests.* (Tech. Rep. No. 26). Urbana: University Illinois, Center for the Study of Reading.

Bailey, J., Brazee, P. E., Chiavaroli, S., Herbeck, J. , Lechner, T., Lewis, D., McKittrick, A., Redwine, L., Reid, K., Robinson, B. & Spear, H. (1988). Problem solving our way to alternative evaluation procedures. *Language Arts, 65,* 364-373.

Bank, A., & Williams, R. C. (1987). *Information systems and school improvement: Inventing the future.* New York: Teachers College Press.

Beady, C., & Slavin, R. (1981). *Making success available to all students in desegregated schools: An experiment.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Bresee, C. W. (1976). On "Grading on the Curve."..*The Clearing House, 5,* 108-110.

Broadfoot, P. (1986). Records of achievement: Achieving a record? *Studies in Educational Evaluation, 12,* 313-323.

Burton, N.W. (1978). Societal standards. *Journal of Educational Measurement, 15,* 263-271.

Butler, R. & Nisan, M. (1986). Effects of no feedback, task- related comments and grades on intrinsic motivation and performance. *Journal of Educational Psychology, 78,* 210-216.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). (pp. 255- 296). New York: Macmillan Publishing Company.

Cohen, E. G. (1986). *Designing groupwork: Strategies for the heterogeneous classroom.* New York: Teachers College Press.

Davis, R. B., & McNight, C. (1976). Conceptual, heuristic, and S-algorithmic approaches in mathematics teaching. *Journal of Children's Mathematical Behavior, 1* (Supplement 1), 271-286.

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist, 34,* 391-401.

Dornbusch, S. M., & Scott, W. R. (1975). *Evaluation and the exercise of authority.* San Francisco: Jossey-Bass.

Dorr-Bremme, D. W. (1983). Assessing students: Teachers' routine practices and reasoning. *Evaluation Comment, 6*(4), 1-12. (ERIC Document Reproduction Service No. ED 242 742).

Doyle, W. (1983). Academic work. *Review of Educational Research, 53,* 159-199.

Ediger, M. (1975). Reporting pupil progress: Alternative to grading. *Educational Leadership, 32,* 265-267.

Gaston, N. (1976). Evaluation in the affective domain. *Journal of Business Education, 52,* 134-136.

Giannangelo, D. M. (1975). Make report cards meaningful. *The Educational Forum*, *39*, 409-415.

Giannangelo, D. M., & Lee, K. Y. (1974). At last: Meaningful report cards. *Phi Delta Kappan*, *55*, 630-631.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist*, *18*, 519-521.

Glass, G.V. (1978). Standards and criteria. *Journal of Educational Measurement*, *15*, 237-261.

Heller, L. (1978). Assessing the process and the product: An alternative to grading. *English Journal*, *6*, 66-69.

Herman, J., & Dorr-Bremme, D. W. (1984). *Testing and assessment in American public schools: Current practices and directions for improvement.* Los Angeles: University of California at Los Angeles, Center for the Study of Evaluation.

Holmes, M. (1978). Evaluating students in the affective domain. *School Guidance Worker*, *33*, 50-58.

Holtz, R. E. (1976). More than a letter grade. *Social Education*, *14*, 23-24.

Ingenkamp, K. (1986). The possible effects of various reporting methods on learning outcomes. *Studies in Educational Evaluation*, *12*, 341-350.

Linn, R. L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, *20*, 179-189.

Lissman, U. & Paetziod, B. (1983). Achievement, feedback, and its effects on pupils -- a quasi-experimental and longitudinal study of two kinds of differential feedback, norm-referenced and criterion-referenced feedback. *Studies in Educational Evaluation*, *9*, 209-222.

Maxfield, B. (1983). *Computer aided profiling: A case study at the North Warwickshire College of Technology and Art.* (ERIC Document Reproduction Service No. ED 237 651).

McPartland, J. M. & McDill, E. L. (1977). Research on crime in schools. In J. M. McPartland & E. L. McDill (Eds.), *Violence in schools: Perspectives, programs, and positions* (pp. 3-33). Lexington, MA: D.C. Heath.

McDill, E. M., Natriello, G. N., & Pallas, A. M. (1985). Raising standards and retaining students. *Review of Educational Research*, *55*, 415-434.

Natriello, G. (1982). *Organizational evaluation systems and student disengagement in secondary schools*. St. Louis, MO: Washington University, Final Report to the National Institute of Education.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*, 155-175 .

Natriello, G., & Dornbusch, S. M. (1984). *Teacher evaluative standards and student effort*. New York: Longman.

Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York: Teachers College Press.

Natriello, G., & McPartland, J. (1987, April). *Adjustments in high school teachers' grading criteria: Accommodation or motivation?* Paper presented at the annual meeting of the American Educational Research Association. Washington, DC.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Parkay, F.W. (1987). The success-oriented curriculum. *The Clearing House, 61*, 66 - 68.

Phoenix Union High School District. (1981). *Instructional manual: English/writing. E.S.E.A. Title One.* (ERIC Document Reproduction Service No. ED 216 069).

Rheinberg, F. (1983). Achievement evaluation: A fundamental difference and its motivational consequences. *Studies in Educational Evaluation, 9*, 185-194.

Rist, R. C. (1970). Student social class and teacher expectations: The Self- fulfilling prophecy in ghetto education. *Harvard Educational Review, 40*, 411-451.

Rosenholtz, S. B. (1977). *The multiple ability classroom: An intervention against the self-fulfilling prophecy.* Unpublished doctoral dissertation, Stanford University.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart, and Winston.

Rudman, H. C. (1987). Testing and teaching: two sides of the same coin? *Studies in Educational Evaluation, 13*, 73-90.

Rudman, H. C., Kelly, J. L., Wanous, D. S., Mehrens, W. A., Clark, C. M., & Porter, A. C. (1980). *Integrating assessment with instruction: A review (1922- 1980).* East Lansing, MI: Michigan State University College of Education, Institute for Research on Teaching.

Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*(4), 455-498.

Shepard, L. A. (1976). *Setting standards and living with them.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Slavin, R. E. (1978). Separating incentives, feedback, and evaluation: Toward a more effective classroom system. *Educational Psychologist, 13*, 97-100.

Solo, L. (1977). What we do because testing doesn't work. *National Elementary Principal, 56*, 63-66.

Stewart, W. J. (1975). A multi-dimensional evaluating-reporting system in the elementary school. *Reading Instruction, 12*, 174-176.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice, 5*(2) 5-17.

Terwilliger, J. G. (1977). Assigning grades -- philosophical issues and practical recommendations. *Journal of Research and Development in Education, 10*, 21-39.

Thorndike, R. L. (1969). Marks and marking systems. In R. L. Ebel (Ed.) *Encyclopedia of educational research* (pp. 759-766). New York: Macmillan.

Walling, D. R. (1975). Designing a "report card" that communicates. *Educational Leadership, 32*, 258-260.

Wilson, S. (1976). You can talk to teachers: Student-teacher relations in an alternative high school. *Teachers College Record, 78*, 77-100.

Wise, R. I., & Newman, B. (1975). The responsibilities of grading. *Educational Leadership, 32*, 253-256.

Withers, G. (1986). Profile reports--A typology and some caveats. *Studies in Educational Evaluation, 12*, 325-334.